

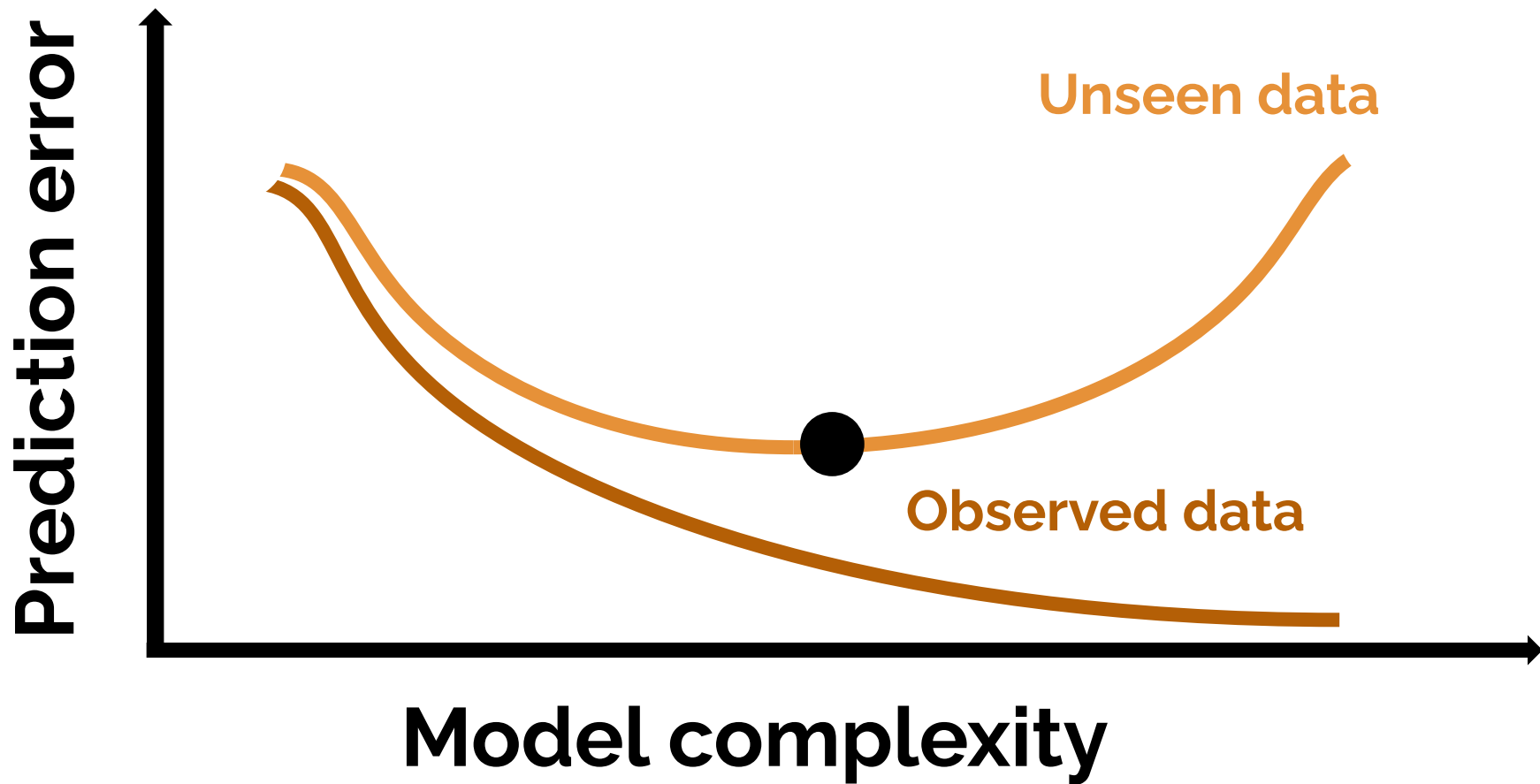
What is Cross Validation?

JS



Central challenge in machine learning

- **Algorithm must perform well on
new inputs, not just training data**



2 Goals of Machine Learning

- **Model selection**
- **Model assessment**

How do you choose a model? 🤔

- Find a model that fits future data best

How do you evaluate a model? 🧐

- Simplest approach: holdout cross-validation

Holdout cross-validation

- Randomly split data into:
 - Training set
 - Testing set

NEVER use test set to train

Holdout cross-validation 🙅

- Fails to use all available data
- If $\frac{1}{2}$ of data is for training, poor hypothesis
- If $\frac{1}{10}$ of data is for testing, poor estimate of actual accuracy

What is the solution?

- **Cross validation**
 - Repeat training and validation on different randomly chosen subsets of original training set

K Fold cross validation

- **Each sample will be used for validation in 1 of the sets**
- **Each sample serves double duty: as training set and validation set**

What is Cross Validation? ✓

JS

